DOCUMENT RESUME

ED 249 231 TM 840 463

AUTHOR Oxford-Carpenter, Rebecca L.; And Others
TITLE Army Job Training Development and Testing Practices Compared to the Instructional Systems Development Model.
INSTITUTION Army Research Inst. for the Behavioral and Social Sciences, Alexandria, Va.
PUB DATE Apr 84
CONTRACT MDA-903-82-C-0531
NOTE 27p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984). Small print in Table 2.
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Adults; Criterion Referenced Tests; Instructional Development; *Instructional Systems; Job Training; *Military Training; *Models; Performance Tests; Test Construction; *Testing; Testing Programs; Test Reliability; Test Validity
IDENTIFIERS *Army; *Instructional Systems Development Model

ABSTRACT

This paper presents an evaluation of Army job training development and testing practices, with a focus on Advanced Individual Testing. Information comes from intensive interviews with school instructors and from observations in the schools. Results indicate that some aspects of the Instructional Systems Development (ISD) model have been implemented less than ideally in Army training installations, while others have been implemented well or are improving. Many testing and training practices are not standardized within or across courses, though courses are almost all lock-step or group-paced at this time. Reliability and validity indices are not calculated by most schools. On the positive side, most tests appear to be criterion-referenced and hands-on, though there is a blurring of distinctions among test types in current Army use. Seven major implications are drawn from these findings. They relate to standardization of training development and test practices, interaction between schools which develop training and tests and schools which instruct, the use of audit trails, statistical properties of tests, the use of hands-on tests, philosophical and instructional conflicts, and the cyclic nature of Army training development and testing practices. (Author/BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED249231

ARMY JOB TRAINING DEVELOPMENT AND TESTING PRACTICES COMPARED TO THE INSTRUCTIONAL SYSTEMS DEVELOPMENT MODEL

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R.L. Oxford-Carpenter

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Rebecca L. Oxford-Carpenter
Advanced Technology

Linda J. Schultz
Helen M. Sperling
Michael G. Rumsey
Army Research Institute

John N. Joyner
Human Resources Research Organization

TM 840 463

Running Head: Army Job Training and Testing

ABSTRACT

This paper presents an evaluation of Army job training development and testing practices with a focus on Advanced Individual Training. Information comes from intensive interviews with school instructors and from observations in the schools. Results indicate that some aspects of the Instructional Systems Development (ISD) model have been implemented less than ideally in Army training installations, while others have been implemented well or are improving. Many testing and training practices are not standardized within or across courses, though courses are almost all lock-step or group-paced at this time. Reliability and validity indices are not calculated by most schools. On the positive side, most tests appear to be criterion-referenced and hands-on, though there is a blurring of distinctions among test types in current Army use. Seven major implications are drawn from these findings.

ARMY JOB TRAINING DEVELOPMENT AND TESTING PRACTICES COMPARED TO THE INSTRUCTIONAL SYSTEMS DEVELOPMENT MODEL

The purpose of this paper is to evaluate Army job training development and testing practices in the framework of Instructional Systems Development (ISD), the model which was mandated to guide Army job training in the last decade. The focus is on measurement and associated issues in Advanced Individual Training (AIT), where soldiers are trained in jobs, known as military occupational specialties (MOS), to which they have been assigned. Measurement in the earlier, more general Basic Training phase and in the later period when soldiers begin actual job duties in their units is beyond the scope of this paper.

Information reported in the current research comes largely from an ongoing, large-scale effort (known as Project A) to improve the Army's selection, classification, and utilization of enlisted personnel. Part of that project involves an evaluation of the Army's current job training measurement system. it is essential to examine this system and to improve it if deficiencies are found. Unless the Army has an effective means of measuring a soldier's progress and proficiency in job training, later classification and utilization of that soldier will be made far more difficult. Individual careers and the effectiveness of the Army as a whole are influenced by the quality of job training and job training measurement available. This is the first time the Army has taken a broad-scope, systematic look at its job training measurement system as part of the larger picture of enlisted personnel selection, classification, and utilization.

General Questions

Evaluation of practices will focus on two general questions: (1) to what extent are training development, test development, and test administration integrated into a coherent system, and (2) what additional information is available concerning test quality?

The first general question can be answered principally on the basis of the extent to which Army training and testing are conducted in a manner consistent with the ISD

philosophy prescribed for them to some extent in Training and Doctrine Command (TRADOC) Regulation 350-7 (Department of the Army, 1982) entitled A Systems Approach to Training, and in the earlier TRADOC Pamphlet 350-30 (Department of the Army, 1975). ISD is the model of job training curriculum design adopted in the 1970s by all four armed services--Army, Navy, Air Force, and Marines. Although in 1984 this model lost support from some high-level Army leaders, for the last decade ISD has held philosophical and operational sway over training development and measurement. It would be difficult to purge ISD philosophy and practice from Army job training, particularly without a clear substitute. Many Army "training watchers" predict that the ISD model may continue in various forms for some time. Others predict that ISD may be officially resurrected, perhaps under a different name or with some alterations, after a laissez-faire period; in other words, the cyclic pattern seen in public school instruction may occur in Army job training. Whatever the case, the ISD model still presents the only relevant framework for reviewing recent training development and related measurement issues in the Army.

ISD presents a process through which training alternatives are selected. Measurement is generally conducted in accordance with the philosophy or model of training employed. The extent to which tests are developed and administered in a manner congruent with ISD philosophy is an issue, therefore, worthy of investigation.

The second general question can be answered through empirical means, chiefly by looking at available reliability and validity data and at practices affecting reliability and validity.

The rest of this paper is divided into six sections: research questions, definitions of testing terms, findings from recent research, methodology of the current investigation, results, and implications for future Army training measurement. We will now turn to the first area of concern, research questions.

Specific Research Questions

There are five specific research questions. The first three, which concern the

extent to which training and testing are integrated into a coherent system, are shown below:

1. To what extent does the training development process follow the ISD model?
2. To what extent are tests developed in a manner consistent with the ISD model?
3. To what extent are tests administered and scored in a manner consistent with the ISD model?

The last two concern additional information on test quality. They are as follows:

4. What information is available on test validity and reliability?
5. What additional information is available on practices of test administration affecting test quality?

Definitions of Testing Terms

Having outlined the specific research questions, we will now define a set of testing terms important to an understanding of job training measurement. It is essential to distinguish between and among types and physical modes of tests.

Test Types

The first important distinction concerns test type, i.e., "norm-referenced tests" (NRTs) vs. "criterion-referenced tests" (CRTs). NRTs are used to assess the achievement of the individual compared to a norm group. Because of the comparative nature of NRTs, they allow the ranking of individuals. CRTs, on the other hand, are not concerned with norm groups or ranking but are instead focused on determine whether an individual has met a prespecified goal, known as a "criterion," "standard," or "cut-off score." NRTs lack the criterion, while the criterion is by definition part of CRTs. If the criterion is set low enough to reflect the minimum competency expected for a given set of tasks, the CRT can be called a "minimum competency test." If the criterion is set higher than minimum competency with the intent of reflecting mastery, the CRT can be termed a "mastery test." The term "proficiency test" is often used to indicate a CRT. Additional information on NRTs and CRTs is found in Angoff (1971), Ebel (1972,

1975), Kriewall (1972), Mehrens and Lehmann (1978), Oxford-Carpenter and Schultz (1983), and Popham and Husek (1969).

Physical Mode of Testing

The second major distinction can be made among physical modes of job training measurement. Some of the key modes include "hands-on" (sometimes called "performance") testing, "paper-and-pencil" (sometimes called "knowledge") testing, and "simulations." Vineberg and Joyner (1983) provided a detailed explanation of the differences between "hands-on" and "paper-and-pencil" tests. To summarize their discussion (pp. 241-248), paper-and-pencil tests have the advantage of efficiency, because they can be used to test certain types of knowledge quickly and inexpensively. However, paper-and-pencil tests have several drawbacks. Such tests often place on the respondent heavy verbal demands that may not be relevant to the job training requirements to be tested. Paper-and-pencil tests are by nature unable to represent all task elements. Artificiality is introduced, because paper-and-pencil tests supply cues not available to an actual job performer and because they require the respondent to select rather than generate the correct response. In contrast, hands-on tests have the potential for greater job fidelity than paper-and-pencil tests. However, hands-on tests are time-consuming, expensive, unable to test many tasks in a given situation due to practical constraints, and vulnerable to non-standardization and unintentional coaching or feedback. Vineberg and Joyner (1983) suggested that knowledge recall is tested best by paper-and-pencil tests, while motor or perceptual tasks are measured most adequately by hands-on tests. Application of a procedure might be amenable to either mode.

A third physical mode of testing is "simulation," which offers more job fidelity than paper-and-pencil measures but less than hands-on measures. Simulation imitates certain environmental or other conditions for the purpose of training, testing, or experimentation. Simulation is less costly than hands-on testing but more expensive than paper-and-pencil testing. Application of rules or principles can be well tested by simulation, according to Vineberg and Joyner (1983).

It is important to realize that NRTs need not be in the paper-and-pencil mode and that CRTs need not be hands-on tests or simulations. Either type of test, NRT or CRT, can be given in any physical mode.

An understanding of the distinctions mentioned here provides a useful background for results of recent research, a discussion of which is in the next section.

Recent Research Findings

Findings of recent research fall into two relevant categories: research on military job training development and research on job training measurement. While these categories overlap to some extent, it is profitable to examine them separately.

Research on Military Job Training Development

The ISD model resulted from two factors: a recognition that the future would demand increasingly large pools of well-trained military personnel in highly technical areas and a reaction against conventional and unsystematic training techniques. ISD is an iteractive, systems approach to training which evolved out of operations analysis concepts of World War II and which was introduced to the Army in 1975. Its 1982 Army revision preserved the ISD procedure while offering greater flexibility to schools for implementation. This revision also called for careful documentation of each ISD phase through "audit trails."

Depending on one's point of view, ISD can be considered a means of designing training that will (1) optimize total system effectiveness, (2) maximize efficiency within the training subsystem, (3) insure that training is relevant to the job, or (4) reflect modern training technology (Vineberg and Joyner, 1980, p. vi). In their evaluation of the current implementation of ISD across the four services, Vineberg and Joyner determined that ISD was at that time not meeting most of these intents for a number of technical and operational reasons, some of which are detailed later.

The five typical phases of the ISD model are analyzing the job (known as "front-end analysis"), developing objectives and tests, specifying learning events/activities,

implementing the instructional management plan, and conducting internal and external evaluations for system revision. A summary of those phases follows.

The job analysis phase determines what constitutes adequate on-the-job performance. Theoretically, it involves assessing the discrepancy between the training that is being given and the training that must be given to satisfy the requirements of the threat, the mission, and the job. Job analysis is accomplished by collecting relevant data about specific, observable, measurable, and finite tasks. Costs of training and task criticality are two factors used in determining which tasks will be selected for training. Construction of accompanying task performance measures, which were formerly called job performance measures in the 1975 TRADOC pamphlet on the model, is a critical element in this phase. These measures are the fundamental basis for control and validation of training. Analyzing existing courses and selecting the appropriate instructional setting are also included in the job analysis phase.

The next phase translates job requirements into learning objectives to be trained and tested. Learning objectives are hierarchically ordered and contain conditions (input), performance behaviors (action), and standards (output). Test item development is also included in this phase. Test purposes include outcome measurement and provision of diagnostic information. Tests can cover either aptitude or acquired knowledge and skills. They include entry tests, pretests, unit posttests, and end-of-course tests. Establishment of criteria or cut-off scores is done to dichotomize trainees into GO and NO-GO categories. In this phase entry behavior is described, and the sequence and structure of the training course are determined.

During the third phase, which focuses on instructional development, learning events and activities are specified, learning guidelines are identified, the instructional management plan and the delivery system are detailed, instructional materials are reviewed and selected, and instruction is developed and then validated.

The fourth phase addresses implementing the instructional management plan and conducting the instruction. The final phase concerns conducting both internal and external evaluations in order to allow for appropriate system revision.

Using detailed interviews of training developers at 33 military organizations and a questionnaire survey of 209 military training agencies, Vineberg and Joyner (1980) found that some of the major components of ISD were not implemented ideally. For example, specification of training objectives was virtually universal across the four services, but the procedures used to identify objectives were highly variable and often unclear. Results of front-end analysis were often ignored in developing training objectives. Also, training objectives were frequently ignored in developing achievement tests. Training methods, media, grouping, and scheduling were not systematically chosen but were almost always imposed by command policy or by such factors as equipment availability or operational schedules. No criteria existed to determine whether training should be revised as a result of external evaluation. These investigators suggested many ways in which ISD could be made more usable for the armed forces, chiefly by making more explicit the elements of each ISD phase and step and the relationships across steps. For some of the implementation deficiencies Vineberg and Joyner recommended that no changes be made at that time due to lack of a cost-effective way to bring about needed change. Clearly, their research showed that ISD is an important move toward making training development a more systematic and rational process, but the actuality falls short of the ideal.

Vineberg and Joyner (1980) warned that although the ISD model emerged about the same time as the national trend toward criterion-referenced testing and associated techniques (individualization, self-pacing, computer-assisted instruction, and computer-managed instruction), ISD should not be considered synonymous with those techniques. ISD is primarily a process through which training alternatives are selected rather than a set of specific training strategies or products. However, in practice ISD clearly implies the use of CRTs because front-end job analysis and development of objectives are designed to determine adequate on-the-job performance--the criterion toward which the training moves.

Research on Military Job Training Measurement

Swezey, Pearlstein, and Ton (1975) conducted research on Army testing practices

in job training schools, training centers, and Army Extension Courses. In interviewing 105 testing and training personnel, these researchers found that the primary use of test results was to evaluate soldier skill performance. Test results were also frequently used to improve training, diagnose trainee weaknesses, and compare trainees. Most measures were either paper-and-pencil tests (47%) or hands-on tests (42%) rather than simulations (8%) or other modes (3%). The investigators found that it was difficult to draw a sharp line between hands-on and paper-and-pencil tests, because some hands-on tests involved paper-and-pencil components, such as making mathematical calculations for aiming artillery. Swezey et al. summarized their findings by saying that job training tests were mostly multiple-choice, paper-and-pencil, and norm-referenced, although they found evidence that interest in and use of CRTs were increasing.

In terms of specific test development practices, Swezey et al. learned that item pools were not often used and that test developers frequently wrote only the precise number of items required for a specific test. No empirical item analysis strategy was used in most cases. There was virtually no consideration of reliability and validity by the test developers interviewed. Lack of standardized test conditions existed in many areas. In short, many basic principles of adequate test development were not being met.

Vineberg and Joyner (1983) made some helpful generalizations about the military job training measurement system as they had known it over a number of years. They stated that job training measurement was the most frequent type of individual performance assessment used in the military. For most jobs, said Vineberg and Joyner, achievement testing in military job training schools was accomplished mainly through paper-and-pencil, multiple-choice tests of information recall. Hands-on testing was used in certain occupations, such as those of rifleman, pilot, and air traffic controller. Tests were used to measure trainees' progress and diagnose deficiencies. Tests were aiso used, though less frequently, to validate selection and classification tests or as a surrogate for job performance. Vineberg and Joyner (1983) also discussed differences between military and industrial testing and characteristics of military proficiency testing programs (Skill Qualification Tests, for example) occurring after initial job training.

We have described findings of relevant research in areas of training development and testing practices. With this basis, we now turn to the methodology used in the current investigation.

Methodology

This section presents the methodology of the current research in two parts: data collection and data analysis.

Data Collection

Structured interviews were held with 12 job training instructors and supervisors in 11 courses, as shown in Table 1. These courses were part of a larger sample of courses representing jobs selected because of their high density and importance to the Army mission. Taken together, the larger sample of jobs covers approximately half of all Army enlisted personnel. While the results of the current investigation do not cover the entire universe of Army jobs, nevertheless these findings are indicative of important trends in Army job training measurement.

INSERT TABLE 1 ABOUT HERE

Interviews lasted approximately one-and-one-half to two hours. Interviewers checked the accuracy of their interpretations with interviewees throughout the discussions and, when necessary, called later for more in-depth information.

Researchers examined training documents, such as Soldier's Manuals, field manuals, and Programs of Instruction (POI). They also took notes on informal observations made on visits to training schools. Finally, they arranged for transmittal of training performance data to a longitudinal data base for future analysis. These data on training performance of soldiers are not reported in the current paper.

Data Analysis

Interview data were content-analyzed by means of response category matrices.

Training documents were also content-analyzed in matrix form according to specific types of activity or task. Informal observations were used for anecdotal information only.

Results

Results reported here are for the 11 courses in which interviews were conducted. The following discussion features five categories of results in response to the five specific research questions listed earlier. The first three categories deal with consistency of training development, test development, and test administration and scoring with the ISD model. The last two categories concern test quality issues.

Consistency of Training Development with ISD Model

The first specific research question concerns the extent to which training development follows the ISD model. Interviews, training document review, and informal observations indicated that the ISD model was indeed being implemented in some form. An important and expectable finding, given the nature of the ISD model as currently implemented in Army training, was that there was little colaboration or communication between the developers of training courses and tests (who are housed mainly in "proponent" schools) and training instructors (who may be housed at other training installations for jobs that are high-density). Instructors generally did not know how or why tests and courses were developed as they were and provided little or no feedback to developers of courses or tests. The job of instructors was to conduct the courses and give the tests rather than to create courses and tests or provide feedback to developers. Of course, the current investigation did not deal with smaller, lower-density jobs, for which training development and actual delivery of training take place at the same installation. For lower-density jobs, which make up a majority of Army jobs but contain a minority of the soldiers, the relationship between trainers and training developers may be closer than for higher-density jobs.

Course content was specific to the job, for the most part. A standard curriculum was available for each course in the form of a POI. Other job-related documents, such

as Soldier's Manuals and Technical Manuals, were also available. These materials were designed and produced by the proponent schools under the guidance of TRADOC. Schools assumed responsibility for implementation and administration of the courses with TRADOC guidance. Emphasis given to different curriculum elements in a course, however, was subject to variation according to operational needs and instructor preferences. Moreover, in some cases alternative course materials selected by an individual school were used in lieu of those recommended by the proponent school. Therefore, course content was not always found to be uniform by MOS.

With respect to instructional mode, all of the courses were group-paced, except for the self-paced 05C (radio teletype operator) course, the mostly self-paced 16S (MANPADS crewman) course, and the lock-step 19E (tank crewman) and 19K (M-1 crewman) courses. Since group-paced and lock-step were virtually indistinguishable delivery modes, the 05C and 16S courses were the only real exceptions--and 16S was scheduled to become group-paced or lock-step in the near future.

Consistency of Test Development with ISD Model

Now we address the second specific research question: To what extent are tests developed in a manner consistent with the ISD model? Relevant results cover such areas as test type, physical mode of testing, test coverage, and item development. Each of these will be discussed in turn as they relate to ISD.

Test type. As noted above, ISD is not synonymous with criterion-referenced testing, but does clearly imply the use of CRTs. Thus, it is noteworthy that all of the hands-on tests were CRTs, because they were scored on a pass-fail, GO/NO-NO basis using specified criteria. It was interesting to note that the CRT emphasis coexisted with the lock-step or group-paced delivery system in current Army job training courses; CRT usage ordinarily goes along with self-pacing in instructional settings. Even more significant was the fact that the current criterion-referenced testing mode represents a shift from the findings of Swezey et al. (1975), who found that NRTs were the typical test variety and that CRTs were only beginning to be used a decade ago. It appears that in the decade since Swezey et al. investigated the Army job training measurement

system, a philosophical and practical change has taken place as reflected in the overwhelming use of CRTs currently.

It should be noted, however, that criterion-referenced scoring is encouraged, if not mandated, by current Army training policy. As a result, instances exist in which NRTs have been converted into CRTs by overlaying a CRT scoring scheme, but with no apparent rationale. For example, a paper-and-pencil knowledge test might be scored on a percentage basis and the percent scores then converted into pass/fail outcomes.

Also, it should be kept in mind that the distinction between NRTs and CRTs becomes blurred in the case of entry-level Army job training, which is provided in AIT. Unlike some civilian education programs, one purpose of which is to establish the relative proficiency of their graduates, entry-level job training in the military has only two major outcomes: graduation and attrition (the latter being rare). Thus, even when AIT tests are given that appear to be norm-referenced due to their scoring scheme, the results of the tests are often used only to sort trainees into those who need remedial training for graduation and those who do not. If no further use is made of information about relative standing, the tests could just as well be considered CRTs.

Physical mode of testing. The primary modes available are hands-on, paper-and-pencil, and simulation. ISD does not dictate a particular mode of testing but does imply that the mode should be systematically related to the learning objective and that the learning objective be systematically related to on-the-job performance, which is often basically of a hands-on nature. Thus, a finding that most or all tests were of the written variety would be a basis for suspicion that the ISD philosophy was not being conscientiously applied. A finding that most or all tests were hands-on would imply that ISD was being followed in relation to job performance and training needs.

It is thus significant that the use of hands-on tests has changed over time. Although Swezey et al. (1975) and Vineberg and Joyner (1983) remarked on the high frequency of paper-and-pencil tests found in Army training courses, the current data showed that most tests now administered for the sampled courses were hands-on rather than paper-and-pencil. (This shift to hands-on testing reflects a recent change in Army policy.) Nine of the 11 courses examined in the current research had at least one

written test, but the written tests were of relatively minor importance compared to the hands-on tests. Two courses, 05C and 71L, had only hands-on tests. Use of hands-on tests may signify a move toward a more valid and relevant assessment mode.

Test coverage of curriculum. ISD also implies comprehensive coverage of the course by the tests. Results here showed that coverage of the course by the tests varied. Although almost everything taught in every course was tested at some time (usually at the end of each unit, module, or "annex"), the end-of-course test covered all the material in only about half of 11 courses. Due to time constraints and the nature of the course material, the other end-of-course tests covered only a sample of the course content. Therefore, the degree of match between test content and course content was subject to variation.

Complying with guidance from TRADOC, many hands-on tests were developed with two or more alternate versions. For most written tests, alternative versions were simply scramblings of the items or the distractors. In some instances, two courses covering the same job were taught quite differently at different training installations, although the same test was used--thus raising the question of content validity of the tests in relation to the course (e.g., 63B). In other courses, such as 94B, one school teaching the course used the proponent school's training and test materials while another school teaching the course used all its own materials. In short, there were many differences across schools and courses in the way course materials and test materials were used.

Item development. Ideal application of the ISD philosophy would perhaps best be represented by a systematic item development procedure in which items were generated from job analysis information. Here, the application of ISD represented something less than the ideal. Items were developed in different ways in different MOS. Some MOS, such as 19E/K, produced the test at the same time as the lesson plan. Guidance for item development for 19E/K came from multiple sources, such as Technical Manuals, Skill Qualification Test Manuals (used for on-the-job testing), and lesson plans. For 94B, on the other hand, most item development was a revision of old test items. In that MOS if an item was consistently answered incorrectly, it was

subject to revision because it was judged to be either faulty or too difficult. Items that were too simple were made more difficult by using a different format or a different conceptual level. Some MOS used item pools, while others did not.

Consistency of Test Administration and Scoring with ISD

The third specific research question concerns the consistency of test administration and scoring with the ISD model. Test administration and scoring results (see Table 2) related to the question of consistency with ISD include number of tests, retesting, challenging, and scoring.

INSERT TABLE 2 ABOUT HERE

Number of tests. ISD does not imply a certain number of tests; it does suggest, however, that testing will be aligned with the training purpose of bringing students to a particular level of competence. Thus, one might expect to find a relatively large number of tests being administered in ISD-driven courses to insure adequate feedback and opportunity for retraining. The results obtained here indicated that the number of tests administered per course varied; however, all the courses carried an end-of-course test. Trainees were expected to achieve competence cumulatively throughout a given course. In all 11 courses, all trainees were expected to pass successively all of the tests given. In two of the 11 courses (19E and 19K), trainees were, on rare occasions, allowed to skip a test but only in cases of medical problems or if prior performance had been outstanding.

Retesting. Retesting was permitted generally in the event of end-of-course failure. The ISD model, which involves training individuals to meet specified objectives, would appear to permit some degree of retesting until the objectives were met, although the retesting limit may not be defined by the model. In most courses, the trainee could be retested immediately or after only a short delay unless retraining was clearly indicated. In the self-paced 05C course, the delay was up to the student. In some courses the trainee who failed a test was retested with another test; in others,

with a different version of the same test; and in others, only only the items missed on the first test. Since the trainers and testers made every effort to see that each trainee passed the course, idiosyncratic treatment was bound to emerge in the testing procedure. Trainees were no longer "recycled" back to the beginning of the course, but there were occasional "new starts" made by trainees who lost time through illness or who were dropped back to a later class. Officially, only two retests were permitted in most courses, but in practice more were sometimes given. If, for example, the failure was slight or based on a misunderstanding, or if the trainee had a superior record to date, additional testing might be allowed.

Challenging. The practice of "challenging" a phase of training by taking the end-of-course test and skipping the training for it if that test is passed has been a feature of self-paced courses and would potentially be allowable under ISD. However, in the only truly self-paced course, O5C, such challenging was rare. Challenges were also allowed for the 63B courses at Ft. Dix and Ft. Jackson, as well as the 94B course at Ft. Dix.

Scoring. To the extent that ISD implies the use of CRT, it also implies the use of a pass-fail scoring systems generally associated with CRT. With respect to determining passing scores, all of the hands-on tests were pass/fail. However, there was no uniformity across courses in criteria used to establish what was "passing." Furthermore, there was no numerical passing score for most hands-on tests. The exception was the final grade in the 16S course--a Progress Index calculated by using observed time as the numerator and expected time as the denominator.

Scores necessary for the individual to pass paper-and-pencil examinations varied from 62.5% to 100%, with no specific decision rule given for the choice of passing score.

The final grade in all but the 16S course was pass-fail or GO/NO-GO for the end-of-course test. In the 16S course, the Progress Index was used as a final grade, with a score of 1.25 or lower required for graduation.

NO-GO on hands-on tests and failing grades on paper-and-pencil tests were treated very differently from course to course and from school to school, just as course tests and instructional emphasis varied.

Availability of Test Quality Information

The fourth specific research question concerns evidence of reliability and validity. Issues in test development which do not bear directly on the extent to which ISD is applied but do reflect on the quality of the tests are reliability and validity. A significant finding was that little or no information existed on the reliability and validity of Army job training measures in the sampled MOS. Because of the lack of concern for statistical properties of tests, the lack of understanding about the importance of reliability and validity of job training measures, and the variable ways in which such measures were developed, it cannot be assumed that these measures would exhibit high reliability and validity if training measure data were analyzed.

Other Factors Related to Test Quality

Other information available on test administration which impacts on test quality includes that on timing and monitoring. These topics were the concern of the final specific research question.

Timing. Time limits on tests were strictly enforced only when timing was a critical factor for successful performance, i.e., the 05C course using International Morse Code and the "ten-second drill" in the 16S course. Instructors reported that in most cases 90% or more of the trainees finished the tasks within the allotted time.

Monitoring. All tests were monitored to prevent cheating. Prompting was almost irrelevant because in many instances the trainee had a manual available during testing, according to informal observations and discussions with instructors.

Summary of Results

Results indicated that certain parts of the ISD model were implemented in a less than ideal way in Army training installations, congruent with earlier findings by Vineberg and Joyner (1980). Other parts of the model--such as the kinds of tests used--were implemented well or were improving.

Courses were almost all lock-step or group-paced, and they appeared to be becoming increasingly so rather than less so. Most tests were reported by instructors to be criterion-referenced, which may represent a trend away from the NRTs of the 1970s and earlier. However, there may be some blurring of distinctions between CRTs and NRTs in current Army usage. Hands-on testing, which represents a move toward potentially more valid measurement, also appeared to dominate the Army job measurement picture in a way that it did not a decade ago. Testing practices were not standardized within courses across installations. In other words, two courses for the same job taught at different installations sometimes used different testing practices. Indices for test reliability and validity were in general not calculated by Army training schools.

Implications

Seven implications arise from the results of the current investigation. They concern standardization, communication, "audit trails," statistical properties of tests, use of different test modes, philosophical and instructional conflicts, and the cyclic nature of training and testing practices.

First, the Army should consider whether there ought to be greater standardization of training development and test practices across training installations for high-density jobs. The question is whether courses should be taught and tested in the same way from school to school. There are, of course, pros and cons for the concept of standardization, and they should be carefully weighed. If a systematic model of instructional development (ISD or another model) were followed more closely, it would allow more standardized Army job training and more accurate job training measurement.

Second, for high-density jobs there needs to be a greater interaction between proponent schools (which develop training and tests) and non-proponent schools (which instruct). Perhaps instructors could teach better if they understood how courses and tests were developed under ISD or another model. Likewise, developers of training courses and tests might benefit from feedback from instructors, who have daily contact with trainees and their learning problems. In the new Army situation, which does not

officially support ISD, instructors may be allowed to develop all their own course materials and tests. While this would eliminate the problem of a gap between instructors and training developers, it would also cause less systematic and more idiosyncratic training and testing.

Third, regardless of the model implemented, Army job training schools need to maintain and use carefully documented "audit trails," as emphasized in the 1982 regulation on the systems approach to training. Such documentation would aid communication and accountability among all parties at every stage in the job training process.

Fourth, greater attention undoubtedly needs to be given to statistical properties of tests. Unless a test can be shown to be reliable and valid, its results may not be meaningful in demonstrating the actual level of soldier performance or achievement.

Fifth, choice of test mode should be related to a number of factors. The growing use of hands-on measurement may increase the relevance and hence the potential validity of some Army job training measurement. Hands-on tests, if well designed and administered, can be a step in the direction of better job training measurement. Simulation-based measures might be more cost-effective than hands-on measures if the proper simulation hardware could be made available. Written tests might also be useful in a limited number of circumstances. Hands-on, simulation-based, and written tests should all be considered for future Army job testing, depending on the nature of the performance to be measured and the fiscal and operational constraints involved.

Sixth, several philosophical and instructional conflicts require examination. The Army should consider the potential conflict inherent in employing lock-step or group-paced courses while using tests that are ostensibly criterion-referenced (and hence slanted toward potential individualization or self-pacing of instruction). Another large conflict that the Army ought to examine lies in the fact that the joint services are developing self-paced, computer-assisted or computer-based instruction (for example, the Triads Computer-Based Instruction Program) at the same time that the Army is reverting to group-paced or lock-step job training. Clarification of philosophical and instructional assumptions and values would be helpful.

Seventh, practices in Army training development and associated testing practices appear to be cyclic. Old practices and philosophies often seem to die out, only to be resurrected in other forms either immediately or later. Persons concerned with military training should watch the current cycle with great interest.

REFERENCES

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), Educational measurement, 2nd ed. (pp. 508-600). Washington, DC: American Council on Education.

Department of the Army. (1975). Interservice procedures for Instructional Systems Development (TRADOC Pam. 350-30). Washington, DC: Author.

Department of the Army. (1982, November). A systems approach to training (TRADOC Reg. 350-7). Washington, DC: Author.

Ebel, R. L. (1972). Essentials of educational measurement. Englewood Cliffs, NJ: Prentice-Hall.

Ebel, R. L. (1975) Educational tests: Valid? biased? useful? Phi Delta Kappan, 57, 83-88.

Kriewall, T. E. (1972, April). Aspects and applications of criterion-referenced tests. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Mehrens, W. A., and Lehmann, I. J. (1973). Measurement and evaluation in education and psychology. New York: Holt, Rinehart and Winston.

Oxford-Carpenter, R. L. and Schultz, L. J. (1983). Reading assessment in the Army (ARI Working Paper RS-WP-83-5). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences.

Popham, W. J. and Husek, T. R. (1969). Implications of criterion-referenced measurement. Journal of Educational Measurement, 6, 1-9.

Swezey, R. W., Pearlstein, R. B, and Ton, W. H. (1975). Criterion-referenced testing: A discussion of theory and practice in the Army (ARI Research Memorandum 75-11). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences.

Vineberg, R., and Joyner, J. N. (1980). Instructional Systems Development (ISD) in the armed services: Methodology and application (Contract No. MDA 903-77-C-0355). Alexandria, VA: Human Resources Research Organization.

Vineberg, and Joyner, J. N. (1983). Performance measurement in the military services. In F. Landy, S. Zedeck, and J. Cleveland (Eds.) Performance measurement and theory. Hillsdale, NJ: Erlbaum, pp. 233-250.

NOTE

This research was conducted under Contract MDA 903-82-C-0531 under the auspices of the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) when the first author worked there. Portions of this paper were presented by the first author at the 1984 annual meeting of the American Educational Research Association in New Orleans. The authors thank Dr. N. Kent Eaton, ARI, and Ed Shepherd, Training and Doctrine Command, for their helpful reviews. Major Ronald Tarr of the U.S. Army Infantry School provided useful ideas. We also thank Mildred Cuevas for production assistance.

TABLE 1. COURSES VISITED

MOS	Title	Location Visited
05C[a]	Radio Teletype Operator[a]	Ft. Gordon, GA
16S	MANPADS Crewman	FT. Bliss, TX
19E	Tank Crewman	Ft. Knox, KY
19K	M-1 Crewman	Ft. Knox, KY
63B	Light-Wheeled Vehicle Mechanic	Ft. Dix, NJ
63B	Light-Wheeled Vehicle Mechanic	Ft. Jackson, SC
64C	Motor Transport Operator	Ft. Dix, NJ
71L	Administrative Specialist	Ft. Jackson, SC
76Y	Unit Supply Specialist	Ft. Jackson, SC
94B	Food Service Specialist	Ft. Dix, NJ
94B	Food Service Specialist	Ft. Jackson, SC

[a]Military occupational specialty (MOS) number and title have recently changed.

Table 2. Test Administration and Scoring Practices in Selected Job Training Courses

Items of Information	05C	16S	19E	19K	63B Dix	63B Jackson	64C	71L	76Y	94B Dix	94B Jackson
No. of tests allowed	3[a]	3	3	3	3	3	3	3	3	3[a]	3
Tests pass to graduate	all	all	all	all	all	all	all	all	all	all	all
Can tests be missed	no	no	yes[c]	yes[b]	no	sometimes	no	no	no	no	no
Same test for all	yes	yes	yes	yes	no	no	no	no	no	no	no
Retests recorded	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Interval before retesting	none	variable	none	none	none	none	retrain	none	24 hr. or less	none	variable
Strictly timed tests	some	one	some	some	all	none	some	none	few	none	none
Trainees finish within time limit	almost all	80%	all	almost all	95%	95%	80%	NA	100%	NA	all
Challenges allowed	yes[c]	no	no	no	yes	yes	no	no	no	yes	no
Are tests monitored	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Is prompting permitted	no[d]	no	no	no	no	no	no[e]	no[f]	no	no	no[c]
Passing score (HO)	NA	PI≤ 1.25	NA	NA	NA	NA	NA	NA	NA	NA	NA
Passing score (P&P)	NA	some 100%	80%	80%	70%	75%	70%	NA	100%	70%	62.5-70%
Final "grade"	pass/fail	PI[g]	pass/fail	pass/fail	pass/fail	pass/fail	pass/fail	pass/fail	pass/fail	pass/fail	pass/fail
Final grade calculation	NA	observed time / expected time	NA	NA	NA	NA	NA	NA	NA	NA	NA

[a]Company commander can change rule. [b]For medical disability or for outstanding performance. [c]Extremely rare. [d]Trainee has manual.
[e]Minor prompting allowed. [f]Only for one test. [g]PI (Progress Index) = Observed time/Expected time.